DarwinBenchmark: Evaluating cohort generation and analytics in OMOP CDM databases

Ioanna Nika[1], Maxim Moniat[1], Guido van Leeuwen[1], Ross Williams[1]
[1]Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands

Background

The *Observational Health Data Sciences and Informatics (OHDSI)* [1] community enables large-scale, reproducible research through the *Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)* [2] and the *Health Analytics Data-to-Evidence Suite (HADES)* [3] that implements standardized analytics on mapped observational health data.

The *OHDSI* [1] network is expanding through initiatives such as *EHDEN* [4] and *DARWIN EU®* [5], with a growing number of databases being mapped to the *OMOP CDM* [2]. Simultaneously, the volume of data within each database is increasing over time, as exemplified in Table 1. In addition, more analytical packages are introduced and are frequently used. The generation of cohorts built in *ATLAS* [6] is also increasing. To enable the early detection of errors and performance issues, before they impact research workflows, benchmarking the performance and reliability of cohort generation and analytics across different data volumes and database systems is necessary.

To respond to these needs, *DarwinBenchmark* was developed to evaluate the performance of cohort generation, analytics, and data sources across the *OMOP CDM* [2] databases. It provides a framework for standardized benchmarking across databases and tools.

Methods

DarwinBenchmark's functionality was defined based on user stories derived from interviews with stakeholders in the *OHDSI* [1] community. During these interviews, key challenges in research code execution, including runtime, package, and database system issues, were discussed. Furthermore, implemented solutions were documented, stakeholder roles were clarified, and opportunities for process improvements were identified.

Currently, the package implements two benchmarking functions: cohort benchmarking and package benchmarking. To visualize and explore the benchmarking results, a Shiny Application is provided.

Cohort benchmarking evaluates the execution of cohort definitions across databases using three cohorts built in *ATLAS* [6]. These cohorts contain frequently occurring *concept ID*s from various domains such as *Conditions, Drugs, Observations, Measurements, Devices* and *Procedures*. Each cohort applies progressively more constraints. Initially, subjects qualify by meeting the cohort's entry requirements. This involves having at least one of the specified concept IDs in any of the relevant tables. In subsequent cohorts, the entry criteria are more selective. At the same time, additional inclusion rules are introduced, requiring subjects to have records with the specified concept IDs across a growing number of domains. These inclusion rules are most complex in the smallest cohort. This is done to investigate how cohort entry requirements and inclusion rules impact performance. Overall, cohort generation success, cohort size, and execution time are tracked.

Package benchmarking runs the specified package's benchmarking function and reports the execution success as well as the runtime of predefined tasks. Supported packages include the following *DARWIN*

EU® packages: *IncidencePrevalence* [7], *CDMConnector* [8], *DrugUtilisation* [9], and *CohortCharacteristics* [10]. *HADES* [3] packages can be integrated into *DarwinBenchmark* by exposing a dedicated benchmarking function.

DarwinBenchmark is implemented in R, licensed under Apache 2.0, and is available on GitHub: https://github.com/darwin-eu/DarwinBenchmark.

Results

Initial benchmarks were run on *Integrated Primary Care Information (IPCI)* [11], a Dutch primary care database mapped to the *OMOP CDM* [2]. Both the cohort benchmark and the package benchmark functions were executed against four versions of the *IPCI* [11] database, covering the period from 2023 to 2025. Benchmarks were run ten times to obtain mean and standard deviation for the runtimes. The infrastructure setup was kept the same across all benchmarks.

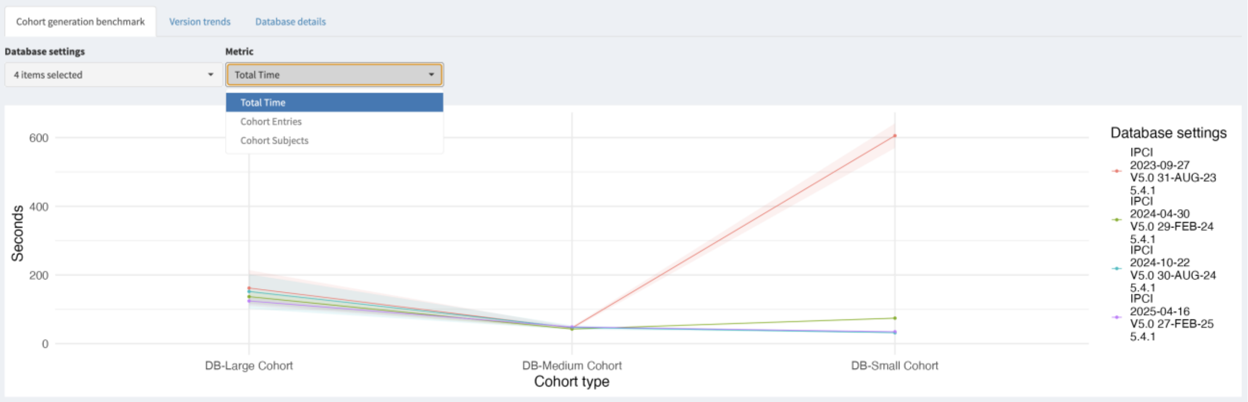
Table 1 shows the number of subjects as well as the total number of relevant records across these database versions. Relevant records include entries from the *Conditions, Drugs, Observations, Measurements, Devices,* and *Procedures* tables. Overall, the number of subjects and records increases with each release as more data is added over time.

The Shiny application offers insights into cohort generation runtimes across databases and their versions. Figure 2 shows that, despite larger data volumes, the 2024-10-22 and 2025-04-16 releases generate the most constrained cohort faster than the 2023-09-27 and 2024-04-30 releases. Runtimes for the remaining cohorts are consistent across versions, with the largest cohort requiring some more time to generate.

Figure 3, presents task runtimes from the benchmarking function of the *CohortCharacteristics* package [10] across different database releases. Findings from the Shiny application indicate that task runtimes remain stable across database releases. The most time-consuming operations involve the *summariseLargeScaleCharacteristics()* function, particularly when applied to the *Measurements* table. The interface also enables comparison of runtimes across different package versions.

Table 1. Number of subjects and total records (sum of entries from the *Conditions, Drugs, Observations, Measurements, Devices*, and *Procedures* tables) across different versions of the *IPCI* [11] database. The number of subjects is rounded to the nearest thousand, and the number of records is rounded to the nearest million.

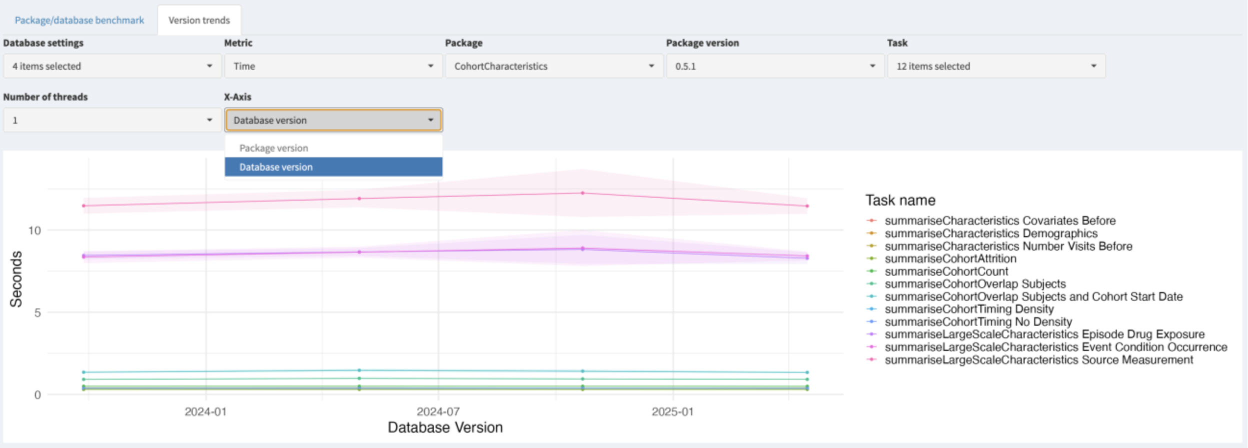
Database Release Date	Number of Subjects (k)	Number of Records (M)
2023-09-27	2817	910
2024-04-30	2870	909
2024-10-22	2955	956
2025-04-16	2985	981


Cohort generation benchmark
Version trends
Database details
Total Time
Cohort Entries
Cohort Subjects
600
400
200
0
Seconds
DB-Large Cohort
DB-Medium Cohort
DB-Small Cohort
Cohort type
Database settings
IPCI 2023-09-27 V5.0 31-AUG-23 5.4.1
IPCI 2024-04-30 V5.0 29-FEB-24 5.4.1
IPCI 2024-10-22 V5.0 30-AUG-24 5.4.1
IPCI 2025-04-16 V5.0 27-FEB-25 5.4.1

Database settings: 4 items selected

Metric: Total Time

Figure 2. User interface for the Cohort Generation Benchmark. Despite larger data volumes, recent releases generate the most constrained cohort faster. Runtimes for other cohorts remain consistent across versions. The largest cohort requires slightly more time to generate.


Package/database benchmark
Version trends
Package version
Database version
10
5
0
Seconds
2024-01
2024-07
2025-01
Database Version
Task name
summariseCharacteristics Covariates Before
summariseCharacteristics Demographics
summariseCharacteristics Number Visits Before
summariseCohortAttrition
summariseCohortCount
summariseCohortOverlap Subjects
summariseCohortOverlap Subjects and Cohort Start Date
summariseCohortTiming Density
summariseCohortTiming No Density
summariseLargeScaleCharacteristics Episode Drug Exposure
summariseLargeScaleCharacteristics Event Condition Occurrence
summariseLargeScaleCharacteristics Source Measurement

Database settings: 4 items selected

Metric: Time

Package: CohortCharacteristics

Package version: 0.5.1

Task: 12 items selected

Number of threads: 1

X-Axis Database version

Figure 3. User interface for Version-Trends from Package Benchmark. Tasks defined in the benchmark function of the *CohortCharacteristics* package [10] are evaluated across multiple versions of the IPCI [11] database. Performance remains stable for all tasks across versions. The most time-consuming task is the use of the *summariseLargeScaleCharacteristics()* function on the *Measurements* table.

Conclusion

DarwinBenchmark addresses the need for systematic benchmarking in *OMOP CDM* databases. It evaluates cohort execution and analytical package performance across databases with varying sizes and infrastructure, to identify runtime and efficiency issues proactively. The package enables reproducible, consistent performance evaluation. It highlights opportunities for improvement for databases and package maintainers to ensure the timely delivery of real-world evidence generated by the *OHDSI* [1] community.

Future work will focus on integrating *HADES* [3] packages into *DarwinBenchmark* and extending benchmarking to include more databases. Finally, additional metrics and statistics will be explored in collaboration with package maintainers and other stakeholders.

References

1. **Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data**

Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574-8.

2. **Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60.**
3. **Schuemie M, Reps J, Black A, Defalco F, Evans L, Fridgeirsson E, et al. Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research. Stud Health Technol Inform. 2024 Jan 25;310:966-70.**
4. **EHDEN Network Expands to 62 Data Partners Across 16 Nations Following Latest Open Call – OHDSI [Internet]. Ohdsi.org. 2023 [cited 2025 Jun 18]. Available from: https://www.ohdsi.org/ohdsi-news-updates/ehden-3rd-open-call/**
5. **The DARWIN EU® Data Network [Internet]. Darwin-eu.org. 2025 [cited 2025 Jun 18]. Available from: https://darwin-eu.org/index.php/data/data-network**
6. **Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. Journal of the American Medical Informatics Association. 2018 Aug 1;25(8):969-75.**
7. **Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. Pharmacoepidemiol Drug Saf. 2024 Jan;33(1):e5717.**
8. **Adam Black, Artem Gorbachev, Edward Burn and Marti Catala Sabate. CDMConnector: Connect to an OMOP Common Data Model [Internet]. Available from: https://github.com/darwin-eu/CDMConnector**
9. **Burkard T, López-Güell K, Gorbachev A, Bellas L, Jödicke AM, Burn E, et al. Calculating daily dose in the Observational Medical Outcomes Partnership Common Data Model. Pharmacoepidemiol Drug Saf. 2024 Jun;33(6):e5809.**
10. **Yuchen Guo, Mike Du, Kim Lopez-Guell, Edward Burn,Nuria Mercade-Besora Marta Alcalde, and Marti Catala. CohortCharacteristics: Summarise and Visualise Characteristics of Patients in the OMOP CDM [Internet]. Available from: https://github.com/darwin-eu/CohortCharacteristics**
11. **de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, et al. Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands. International Journal of Epidemiology. 2022 Dec 13;51(6):e314-e323.**